

No. 2005–06

**BIRTH SPACING AND NEONATAL MORTALITY IN INDIA:  
DYNAMICS, FRAILTY AND FECUNDITY**

By Sonia Bahlotra, Arthur van Soest

January 2005

ISSN 0924-7815

# WORKING P A P E R

---

## Birth Spacing and Neonatal Mortality in India: Dynamics, Frailty and Fecundity

SONIA BHALOTRA  
ARTHUR VAN SOEST

WR-219

January 2005

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.



LABOR AND POPULATION

# **Birth Spacing and Neonatal Mortality in India: Dynamics, Frailty and Fecundity\***

**Sonia Bhalotra, University of Bristol,**

**Arthur van Soest, RAND and Tilburg University**

10 January 2005

## **Abstract**

A dynamic panel data model of neonatal mortality and birth spacing is analyzed, accounting for causal effects of birth spacing on subsequent mortality and of mortality on the next birth interval, while controlling for unobserved heterogeneity in mortality (frailty) and birth spacing (fecundity). The model is estimated using micro data on about 29000 children of 6700 Indian mothers, for whom a complete retrospective record of fertility and child mortality is available. Information on sterilization is used to identify an equation for completion of family formation that is needed to account for right-censoring in the data. We find clear evidence of frailty, fecundity, and causal effects of birth spacing on mortality and *vice versa*, but find that birth interval effects can explain only a limited share of the correlation between neonatal mortality of successive children in a family.

Key words: fertility, birth spacing, childhood mortality, health, dynamic panel data models, siblings

JEL codes: I12, J13, C33

Contact: [s.bhalotra@bristol.ac.uk](mailto:s.bhalotra@bristol.ac.uk), [vansoest@rand.org](mailto:vansoest@rand.org).

---

\* We are very grateful to Wiji Arulampalam for helpful discussions, and to Mike Veall and participants of the RAND Labor and Population seminar for useful comments.

# **Birth Spacing and Neonatal Mortality in India: Dynamics, Frailty and Fecundity**

Sonia Bhalotra and Arthur van Soest

## **1 Introduction**

In developing countries, high fertility is closely related to high levels of childhood mortality. Understanding the way in which family behaviour shapes this relation is crucial to understanding the demographic transition<sup>1</sup> that has historically preceded economic growth. Moreover, the avoidance of child death is, probably one of the most significant aspects of human progress, while sustained reductions in fertility have dramatic implications for the economic independence of women.

Time series analyses of historical data for today's industrialized countries suggest that a marked decline in mortality preceded the decline in fertility (see Mattheisen and McCann 1978, Wolpin 1997), and a similar tendency has been observed in recent aggregate data for sub-Saharan African countries (e.g. Nyarko et al 2003). At the same time, cross-sectional studies using household survey data have produced considerable evidence of the reverse direction of causation, namely that high fertility, associated with close birth spacing, causes an increase in mortality risk within families (e.g. Cleland and Sathar 1984, Curtis *et al* 1993). In families with multiple children, it is easy to see that there is in fact a recursive bi-causal relation of these variables, and that this merits a panel data analysis. The death of a child has been found to result in a shorter interval to the next birth, which may be explained in terms of volitional replacement (see Preston 1985) or else by the fact that the mother stops breastfeeding and, thereby, is able to conceive the next child sooner than otherwise (e.g. Bongaarts and Potter 1983, Chen *et al* 1974). The short birth interval, in turn, results in an elevation of the mortality risk of the *next* child in

---

<sup>1</sup> This refers to the transition from high birth and death rates to low birth and death rates which, in the history of today's industrialized countries, was systematically associated with changes in economic and population growth. For recent theoretical analysis of this relation, see Galor and Weil (2000). Historical analyses of this relation have emphasised timing, arguing that the decline in mortality preceded the decline in fertility, thereby raising issues of causality that have excited attention (see Ben-Porath 1976 for example). We argue below that despite the long-standing interest in this issue, the literature is scarce in a complete micro-data analysis of the inter-relations of these variables.

the family, for example, because the mother has not recuperated physiologically from the previous birth (e.g. DaVanzo and Pebley 1993, Scrimshaw 1996). The latter effect is likely to be especially pronounced amongst poor women who need longer to replenish stocks of nutrients like calcium and iron that are needed to support a healthy pregnancy.

This paper produces estimates of the causal effect of birth interval length on subsequent mortality risk, *and* of mortality on the subsequent birth interval length, exploiting the natural sequencing of these processes to achieve identification. Since the birth-interval data are right-censored, these equations are jointly estimated with an equation for the probability that fertility is incomplete at the survey date, using information on sterilization to aid identification. The estimation allows for endowments, unobservable by the econometrician but known to the parent, and for the agency of the parent in influencing outcomes. Health endowments are referred to as *frailty*. Modeling this term allows for the fact that children of the same mother have correlated mortality risks because of shared genetic or environmental factors. We also allow for inter-family unobserved heterogeneity in the birth spacing and fertility equations (henceforth referred to as *fecundity*), and for this to be correlated with frailty. Thus for example, in our model women who are more careful about contraception may also be more careful in maintaining the health of their children. Ignoring unobserved heterogeneity will not only give biased estimates of the dynamics of each process (see Heckman 1981, Hyslop 1999) but may also bias estimates of the causal effect of each of these variables on the other (see Alessie *et al* 2004). The model is dynamic by virtue of including the mortality status of the sibling preceding the index child. In other words, while mothers describe the cross-sectional dimension of the data, child birth-order creates a time dimension. Allowing unobserved heterogeneity in a dynamic model raises the “initial conditions problem” (e.g. Heckman 1981). This is addressed by exploiting information on the first-born child of every mother, which represents the genuine start of the stochastic process of interest. Allowing for right-censoring and for initial conditions, we have a four-equation model, and this is estimated by (smooth) simulated maximum likelihood.

The paper contributes to the literature in economics and demography. Close relatives of this paper are the models of child health and mortality estimated in Rosenzweig and Schultz (1983a,b) and Rosenzweig and Wolpin (1988, 1995), in which

birth spacing and/or maternal age at birth appear as endogenous inputs (also see the survey in Wolpin 1997). It augments the evidence from these studies in providing estimates of the causal effect of birth spacing on mortality risk, which is of direct relevance to family planning policies. It differs in its modeling approach, described in section 3 below. A further difference of this paper from these studies is that it simultaneously provides estimates of birth spacing and fertility as a function of previous mortality. This is relevant to understanding the overall responsiveness of reproductive behaviour to changes in childhood mortality, which is thought to drive the demographic transition (e.g. Wolpin 1997, Nyarko *et al* 2003) and, thereby, to influence rates of population and economic growth, and the health of surviving children. Previous estimates of this effect have tended to be obtained under the untenable assumption that parents have no influence on the survival chances of their offspring (e.g., Ben-Porath 1976; see Wolpin 1997, Cigno 1998).<sup>2</sup>

Another contribution of this paper is that it introduces lagged mortality (i.e. the survival status of the previous child) in the model, and thereby produces estimates of “state dependence” in childhood mortality within families, and the extent to which this can be explained by birth spacing. State dependence refers to the causal effect of the death of a child on the risk of death of his or her succeeding sibling.<sup>3,4</sup> There is relatively little previous research on state-dependence type effects in analysis of sibling data, although sibling correlations in outcomes have been widely studied (e.g. Solon *et al* 1991). Identification of the (endogenous) effect of an outcome for one individual on the outcome of a “proximate” individual (in this paper, a sibling) is, in general, a challenge

---

<sup>2</sup> Exceptions are Olsen (1980) and Olsen and Wolpin (1983), both of which analyse the response of the number of births to child mortality, allowing it to be correlated with fertility. The approach taken in the current analysis is different in that we use a dynamic panel data framework and also provide estimates of the causal effect of mortality on birth spacing.

<sup>3</sup> State dependence is an expression that has been used in other statistical applications. For example, state dependence in unemployment refers to the causal effect of (individual) unemployment in one period on the risk of unemployment in the following period. Several studies have attempted to disentangle state dependence and unobserved heterogeneity in seeking an explanation of unemployment persistence (e.g. Heckman 1981).

<sup>4</sup> Defining a state as a realisation of a stochastic process, *state dependence* as used here captures the idea that the mortality risk facing a child is dependent upon the state (died in the neonatal period or not) revealed for the previous child in the family. Since time is implicit in the sequencing of children, models that include the previous child’s survival status are analogous to dynamic models.

(e.g. Manski 1993). Here, this is resolved by the natural sequencing of siblings and the fact that, after controlling for heterogeneity, the effects of predetermined variables can be interpreted as causal.

This paper also relates to a larger literature in economics, concerned with the manner in which the allocational decisions of parents affect the quality of their children. Numerous studies have attempted to estimate the effect of schooling, a parental input, on child quality, while allowing for unobserved endowments or ability (e.g. Griliches 1979). This is similar to the problem in the current analysis of allowing endogeneity in inputs to health, given endowment heterogeneity.

Previous demographic research provides estimates of some of the main effects analysed in this paper, although not in a unified framework: for example, see Curtis *et al* (1993) or Madise and Diamond (1995) for analysis of the effects of birth-spacing on mortality, and Zenger (1993) or Frankenberg (1998) for analysis of the effects of mortality on birth-spacing. The limitation of these studies is that their estimates cannot be given a causal interpretation (also see Moffitt 2003), and overcoming this is our main contribution to demographic research. This paper also contributes to the recent literature on sibling death clustering (e.g. Guo 1993, Zenger 1993, Curtis *et al* 1993, Sastry 1997a,b) by emphasizing the distinction between between-family heterogeneity and causal mechanisms such as birth-spacing, learning or maternal depression that operate within families. In this, it takes forward the work of Arulampalam and Bhalotra (2004a,b), who explore the role of state dependence in explaining death clustering, but who do not simultaneously analyse birth spacing.

Some of the interesting questions that our model is able to address are as follows. When parents lose a child, do they conceive the next child sooner than they otherwise would? To what extent do short birth intervals increase the risk of early childhood death? Why are child deaths concentrated in some families? To what extent does this reflect frailty shared amongst siblings, for example, on account of genetic traits, as opposed to causal effects running from the actual event (experience) of death of a child to a higher death risk for the next child in the family? To what extent are the latter, state-dependence type effects mediated by short birth intervals? How do the main endogenous variables

(mortality, birth spacing, fertility) vary with maternal age at birth?<sup>5</sup> How do they vary with parental education, religion, caste, birth-order and gender? Allowing for these sorts of observable characteristics, what is the estimated trend reduction in mortality and fertility (and the estimated shortening or lengthening of birth intervals, as the case may be) over the three decades that the available data span? Overall, the research will throw new light on the dynamics of family formation, focusing on birth and death (rather than on marriage and divorce, the focus in recent research using data from OECD countries-e.g. Akerlof 1998, Cherlin 1990).

Our main findings are summarized here. A neonatal death shortens the subsequent birth interval by about 20 per cent. This, in turn, raises the probability that the next child in the family dies in the neonatal period by about 1 percentage-point. With birth interval length held constant, there is an additional risk-raising effect of the preceding sibling's mortality of about 4.3 percentage-points. So the total impact of a neonatal death on the risk of subsequent neonatal death in the same family is estimated at a remarkable 5.2 percentage-points, and this is after all sources of inter-family heterogeneity held constant. It is notable that birth-interval-related mechanisms can explain only a fifth of within-family state dependence in mortality. This suggests a role for other factors, identification of which is an important avenue for further research. One possible mechanism is maternal depression, which has been neglected in the antecedent demographic literature. For instance, if a mother who has lost a child has not recovered psychologically by the time she conceives her next child then this may create physiological effects that make this child more vulnerable both in the womb and after birth.<sup>6</sup>

There is clear evidence of both frailty and fecundity, although there is no evidence that they are correlated. We find that neglecting to allow for frailty and fecundity biases upwards the effect of lagged mortality on mortality of the index child, but that it has no significant effect on the estimated impact of birth interval length on mortality risk, or on the impact of mortality on birth interval length. There is a significantly negative correlation between the fecundity terms in the equations for birth spacing and for

---

<sup>5</sup> Maternal age at birth is also endogenous once birth intervals are endogenous. This is discussed further in section 3.

<sup>6</sup> Steer *et al* (1992), for example, report evidence that depression can cause adverse pregnancy outcomes.



continued fertility, implying that mothers who tend to have shorter birth intervals also tend to have more children, even if age, calendar year, and other characteristics are held constant. We find clear evidence of geographic cluster effects. Although incorporating these effects increases the standard errors, it does not change the main findings.

The next section describes the data and the endogenous variables. The econometric model is detailed in section 3. The main results are reported in section 4, where we also report some specification checks. Section 5 concludes.

## **2 Data & Descriptive Statistics**

### **2.1 The Data**

The data are from the second round of the National Family Health Survey of India (NFHS-II) which recorded complete fertility histories for ever-married women aged 15-49 in 1998-99, including the time and incidence of child deaths.<sup>7</sup> The sample we use, for the state of Uttar Pradesh (UP), contains 28,668 live births of 6716 mothers, that occurred between 1963 and 1998.<sup>8</sup> UP is the largest Indian state and, in the year 2000, it contained 17.1% of the country's population (approximately 165 million people). It has social and demographic indicators that put it well below the Indian average (see Drèze and Sen 1997).

The incidence of neonatal death over the sample period in UP was 7.39% (7% in the sample used), compared with an all-India average of 5.21%.<sup>9</sup> The percent of birth intervals in the sample that are shorter than 18 months is 18.1, and the percent that are 18-23 months long is 18.6. The mean number of births per mother is 5.32, the median number is 5, the maximum is 14, and as many as 38.3% of women have 5 or more children. The average age of mothers at first birth is 18.4.

---

<sup>7</sup> For details on sampling strategy and context, see IIPS and ORC Macro (2000).

<sup>8</sup> The original sample contains 29,937 births from 7,297 mothers. In line with the demographic literature on mortality, we drop all mothers who have at least one multiple birth. Children of a multiple birth face hugely higher odds of dying, other things equal. Including multiple births would complicate the relation of mortality and birth intervals that is of interest in this paper.

<sup>9</sup> These figures are averages over the data sample. As this contains retrospective data, it includes children born across almost four decades, 1963-1998. Although we do not have recent figures for neonatal deaths, the infant death rate in India is estimated to have been 6.7% in 2001 (UNDP 2003), while the all-India average of the infant mortality rate in our sample is 8.2%. In our all-India sample, 63.4% of infant deaths occurred in the neonatal period.

Female sterilization is the predominant form of contraception in India. At the time of the survey, 22.5% of women were sterilized, information that we use in estimation of the model. Of the women who report sterilization, 59% were sterilized at a parity greater than or equal to 5. In the five years before the survey, 54.4% of women had never used any method of contraception.

The Indian National Health Survey used in this paper is one of a family of about 70 Demographic and Health Surveys (see [www.measuredhs.com](http://www.measuredhs.com)). The methods used in this paper are therefore immediately applicable to a vast array of countries with different profiles of the structural processes. For instance, persistently high fertility and childhood mortality are a greater problem in many African countries than in India. The analysis in this paper could fruitfully be applied to analyze the extent to which the African problem reflects a “demographic trap”, described by the inter-dependence of mortality and fertility.

## **2.2 The Endogenous Variables**

Means and standard deviations of all variables used in the analysis are in Appendix Table 1, where the variables are also defined. The focus in this paper is on neonatal mortality, or death in the first month of life. This assists the statistical modeling since it means that we can be sure that if the preceding sibling died, then this event occurred before the birth of the index child. In other words, lagged mortality is always a predetermined variable in the birth interval equation. Previous research confirms that the association of birth spacing and mortality is strongest in the neonatal period (e.g. Cleland and Sathar 1984, Zenger 1993, Frankenberg 1998, Nyarko et al 2003). Although the focus of policy is on reduction of under-5 mortality (see UNDP 2003), 46.2% of under-5 deaths in India are neonatal (46.4% in UP; figures from the NFHS data), and this proportion has been increasing over time (e.g. World Bank 2004). This is consistent with socio-economic development and “nurture” having a greater impact on survival chances for older children, with biology (“nature”) weighing more heavily in the causes of neonatal death. For this same reason, gender differences in mortality risk are smaller in the neonatal than in the post-neonatal period. As a result, data on boys and girls are pooled in the analysis. Note that, to the extent that the health technology varies with child

age (e.g. Wolpin 1997, p.525), it is less restrictive to estimate models for neonatal mortality separately from models for later childhood mortality than it is to group deaths in the under-5 band.

The birth interval is the interval between reported dates of birth, rather than the inter-conception interval. As a result, measured birth intervals will be shorter on account of premature births (e.g., Gribble 1993). This is investigated in section 4, by dropping mothers with at least one birth interval less than 9 months. A further potential problem is that birth intervals, as measured, will be longer on account of miscarriage or stillbirth (e.g. Madise and Diamond 1995). We do not have reliable estimates of the extent of miscarriage or stillbirth in the data and are therefore unable to assess the impact of this problem.<sup>10</sup>

Figure 1 is a non-parametric regression of the (unconditional) predicted probability of neonatal death as a function of the logarithm of the preceding birth interval. This is seen to decline monotonically. At short birth intervals, not only is the probability of neonatal death highest, but also the gains from an additional month's spacing are largest. Figure 2 plots the kernel density functions of the birth interval for two sub-samples of the data, selected according to whether or not the previous child in the family survived the neonatal period. It shows that the birth interval distribution for the case where the preceding child has died lies to the left of the other. The median birth interval is 22.5 months after a neonatal death and 27 months when the previous sibling has survived the neonatal period (the corresponding means are 24.6 and 30.9 months). The raw data thus exhibit the patterns that we are seeking to quantify.

### 3 The Model

The model has a recursive dynamic structure: the unobservable propensity of neonatal mortality risk depends upon previous mortality in the family (and, thereby, on lagged inputs to child health) and on the preceding birth interval, while the birth interval,

---

<sup>10</sup> Ignoring miscarriage and stillbirth may lead to under-estimation of the mortality-raising effect of short birth intervals in equation (1) below if women who have these problems also tend to produce weaker live births, since then falsely long intervals will be associated with higher mortality. However, this bias may be expected to be small once we control for mother-specific frailty and fecundity.

in turn, depends upon the mortality status of the preceding sibling. Both processes are conditioned on other predetermined or exogenous factors. The effects of birth intervals on mortality and *vice versa* are identified by timing, not by imposing exclusion restrictions on the exogenous variables. The mortality equation can be regarded as a health production function in which the birth interval is an endogenous input (as in Rosenzweig and Schultz 1983a,b). The birth spacing equation is an input equation but, at the same time, it describes an outcome that depends upon both tastes and technology. These two equations are estimated jointly with an equation for continued fertility that accounts for right-censoring of the birth interval, and an equation for mortality risk of the first-born child, that addresses the initial conditions problem.

The econometric model is an extension of the univariate model of Heckman (1981) and Hyslop (1999), and is broadly similar to the bivariate discrete choice model of Alessie *et al* (2004) although, here, the second equation (for birth interval length) is continuous rather than discrete, and subject to right censoring. The way in which the initial conditions problem is addressed is also different in this paper. To take account of sampling design, we introduce a random effect at the community level. The model is estimated by simulated maximum likelihood.

As discussed in section 1, we are not aware of any previous research that provides estimates of causal effects of birth spacing on mortality and of mortality on birth spacing, but there is related previous research on child mortality in which endogeneity of birth spacing has been allowed. For example, Bhargava (2003) argues that the endogeneity of birth spacing is taken care of by controlling for the survival status of older siblings. This, in turn, is instrumented using household possessions and number of previous births, taking birth history to be exogenous. Rosenzweig and Schultz (1983b) directly instrument birth spacing using incomes and prices. However, as discussed in the more recent papers of Rosenzweig and Wolpin (1988, 1995), the implied exclusion restrictions typically do not hold. The more recent work of Rosenzweig and Wolpin (1988, 1995) uses sibling differences to eliminate the mother-specific endowment. In order to further allow for differences across siblings in frailty, they instrument inputs in the differenced equation using “lagged” inputs from older siblings and parental characteristics. The econometric strategy in this paper is similar insofar as it relies upon information restrictions associated

with the sequencing of births. Our use of random effects as opposed to fixed effects at the mother-level is justified below (in section 3.1).

Let there be  $n_i$  children in family  $i$ .  $M_{ij}$  denotes an indicator variable with value 1 if child  $j$  in family  $i$  suffers neonatal death, and 0 otherwise.  $B_{ij}$  is the log of the length of the interval between the birth of child  $j-1$  and child  $j$  in family  $i$ .<sup>11</sup> In other words,  $B_{ij}$  refers to the interval closed by the birth of child  $j$ . As it is the preceding birth interval for child  $j$ , it is, by definition, predetermined with respect to  $M_{ij}$ . The rest of this section describes each of the four equations in the model, and explains the estimation procedure.

### 3.1 Neonatal Mortality

For child  $j$  ( $j=2, \dots, n_i$ ) in family  $i$  ( $i=1, 2, \dots, N$ ), the equation for neonatal mortality is

$$(1) \quad M_{ij}^* = g(\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{ij}, M_{i1}, \dots, M_{i,j-1}, B_{i2}, \dots, B_{ij}; \theta_m) + \alpha_{mi} + u_{mij};$$

$$M_{ij}=1 \text{ if } M_{ij}^* > 0 \text{ and } M_{ij}=0 \text{ if } M_{ij}^* < 0$$

In order to explain the assumptions needed for consistent estimation, it is initially written in a general form. Here  $\alpha_{mi}$  is family (or mother)<sup>12</sup> specific unobserved heterogeneity, reflecting “*frailty*” from genetic sources (e.g maternal propensities to low birth weight and prematurity) as well as from environmental factors and child-care behaviours. As emphasized in Rosenzweig and Wolpin (1988), the fact that endogenous inputs like breastfeeding are not explicitly incorporated implies that the estimated family-effect will reflect not only inter-family heterogeneity in endowments but also any inter-family heterogeneity in preferences. It is assumed that  $\alpha_{mi}$  is known to the family, though not observed by the econometrician. The error term  $u_{mij}$  varies over mothers as well as children. It is revealed at the birth of child  $j$  and we assume that it does not influence parental inputs to child  $j$  in the one month of life during which parental choices can influence neonatal mortality risk. However, we allow  $u_{mij-1}$  to influence parental inputs into child  $j$  through past mortality in the family,  $M_{ij-1}$ .

---

<sup>11</sup> The logarithm of the birth interval is used as this has a more normal distribution than the level in months.

The vectors  $\mathbf{x}_i$ ,  $\mathbf{x}_{i1}$ , and  $\mathbf{x}_{ij}$  are exogenous explanatory variables, partitioned into variables that vary over children ( $\mathbf{x}_{ij}, j=2, \dots, n$ ), that are specific to the first child ( $\mathbf{x}_{i1}$ ) and that do not vary over children ( $\mathbf{x}_i$ ). The vector of unknown parameters is denoted by  $\theta_m$ . The variables  $M_{i1}, \dots, M_{i,j-1}, B_{i2}, \dots, B_{ij}$  are predetermined, i.e., realized at or before the birth of child  $j$ . For the function  $g$ , we will use a linear specification in  $\mathbf{x}_i, \mathbf{x}_{ij}, M_{i,j-1}, B_{ij}$ , and also include quadratic terms in the year of birth of the child, and in the age of the mother at birth, both of which are functions of  $\mathbf{x}_{i1}$  and  $B_{i2}, \dots, B_{ij}$ .<sup>13</sup> Since the age of the mother at birth of child  $j$  depends upon her age at birth of child  $j-1$  and the length of the intervening birth interval,  $B_{ij}$ , it is clear from recursivity of the model that maternal age at birth of  $j$  can be expressed as a function of maternal age at first birth (in  $\mathbf{x}_{i1}$ ) and the history of birth intervals up until that date ( $B_{i2}, \dots, B_{ij}$ ). Thus, by allowing for the endogeneity of birth intervals and conditioning on  $\mathbf{x}_{i1}$ , we are allowing for the endogeneity of maternal age. Since the data used include births that occurred across a span of about 30 years, a quadratic in the year of birth of the child is included to capture any technological change. This, like maternal age, is a function of the year of birth of the first child (assumed exogenous, and in  $\mathbf{x}_{i1}$ ), and the history of previous birth intervals of the mother.

We expect a negative effect of  $B_{ij}$  on  $M_{ij}$  in view of previous evidence (e.g. Cleland and Sathar 1984, Curtis *et al* 1991, Sastry 1997a,b), consistent with the hypothesis of maternal depletion indicated in section 1, and also with competition amongst closely-spaced siblings (e.g. Zenger 1993). The effect of lagged mortality,  $M_{i,j-1}$  on  $M_{ij}$  may be negative if learning effects dominate, or positive if there is a strong role for factors such as maternal depression (indicated in section 1). The first-order Markov assumption implicit in (1) is justified by consideration of the mechanisms that may drive state dependence (that is, a causal effect of  $M_{i,j-1}$  on  $M_{ij}$ ): see Zenger (1993).

We assume that  $\mathbf{x}_i, \mathbf{x}_{i1}$ , and  $\mathbf{x}_{ij}$  are independent of  $\alpha_{mi}$  and  $u_{mij}$ . Mean independence of  $(\mathbf{x}_i, \mathbf{x}_{i1})$  and  $\alpha_{mi}$  is the usual assumption in a random effects model, needed for identification; the conditional mean of  $\alpha_{mi}$  given  $\mathbf{x}_i$  and  $\mathbf{x}_{i1}$  is subsumed in  $g$ . In  $\mathbf{x}_i$ , we include variables reflecting education levels of the mother and father, and ethnicity and

---

<sup>12</sup> Re-marriage (and re-partnering) amongst Indian women is rare enough that it is reasonable to use “mother” interchangeably with “family”.

religion dummies. In  $\mathbf{x}_{it}$  we additionally include calendar year and age of mother at first birth.

A potential drawback of random effects models as compared with fixed effects models is the assumption that the “*time-varying*” (in this context, varying across siblings and, thereby, implicitly over time) regressors  $\mathbf{x}_{ij}$  are assumed to be independent of the individual effects  $\alpha_{mi}$ . In our case, however, the only variables included in  $\mathbf{x}_{ij}$  are child gender and birth-order. Since there seems to be no reason why these should be correlated with mother-level frailty, the independence assumption would seem unproblematic in this model.

### 3.2 Birth Spacing

The equation for the log length of the birth interval is specified in a similar way to the mortality equation:

$$(2) \quad B_{ij} = h(\mathbf{x}_i, \mathbf{x}_{it}, \mathbf{x}_{ij-1}, M_{i1}, \dots, M_{ij-1}, B_{i2}, \dots, B_{ij-1}; \theta_b) + \alpha_{bi} + u_{bij}.$$

The family specific effect in the birth spacing equation,  $\alpha_{bi}$ , is referred to as “*fecundity*” though it will include not only biological fecundity but also any other sources of persistent inter-family heterogeneity that are unobserved, such as the motivation of the mother to engage in market activity (e.g. Heckman, Hotz and Walker 1985). A causal effect of mortality of child  $j-1$  on the birth interval to child  $j$  is allowed through  $M_{ij-1}$ . Past death shocks are, in this way, allowed to influence current behaviour. We include  $\mathbf{x}_{ij-1}$  since the gender of the previous child ( $j-1$ ) may have an effect on the interval to the birth of child  $j$ . The function  $h$  is specified as a linear combination of  $\mathbf{x}_i$ ,  $\mathbf{x}_{ij-1}$ ,  $M_{ij-1}$ , and the calendar year and age of the mother at the time of the birth of child  $j-1$  and their squares. As discussed in section 3.1, the calendar year or year of birth of the child, and maternal age at birth of the child are functions of  $\mathbf{x}_{it}$  and  $B_{i2}, \dots, B_{ij-1}$ . Biomedical and demographic research provide no clear argument for a causal effect of  $B_{ij-1}$  on  $B_{ij}$ , conditional on  $\alpha_{bi}$ , so

---

<sup>13</sup> We experimented with interactions and squares of other terms but found no significant improvement.

we do not allow for this.<sup>14</sup> The assumptions concerning family specific effects and error terms,  $u_{bij}$ , are similar to those for equation (1). We assume that  $\mathbf{x}_i$ ,  $\mathbf{x}_{il}$ , and  $\mathbf{x}_{ij}$  are independent of  $\alpha_{bi}$  and  $u_{bij}$  and that  $u_{bij}$  is independent of the past.

We allow for correlation between the unobserved heterogeneity terms  $\alpha_{bi}$  and  $\alpha_{mi}$  in equations (1) and (2). This allows an alternative, non-causal explanation for the correlation between birth interval lengths and mortality in the raw data. It also accounts for the potential endogeneity of the preceding birth interval in equation (1), which, although predetermined, may be correlated with frailty,  $\alpha_{mi}$ . For example, parents with weak endowments may choose shorter birth intervals in order to meet their target number of children in a given time. Similarly, our model allows  $M_{ij-1}$  in equation (2) to be correlated with family-level fecundity,  $\alpha_{bi}$ .

The distribution of the family effects  $(\alpha_{mi}, \alpha_{bi})$  is assumed to be bivariate normal with mean zero, variances  $\sigma_m^2$ ,  $\sigma_b^2$ , and covariance  $\sigma_m\sigma_b\rho_\alpha$ . The child-specific error terms  $u_{mij}$  and  $u_{bij}$  are assumed to be independent of  $\alpha_{mi}$  and  $\alpha_{fi}$  and normally distributed with mean zero. Without loss of generality, the variance of  $u_{mij}$  is set to 1.

### 3.3 The Initial Conditions Problem

“Lagged” mortality,  $M_{ij-1}$ , is endogenous in equation (1) by virtue of being correlated with frailty,  $\alpha_{mi}$ . This creates the initial conditions problem commonly encountered in analysis of dynamic models with unobserved heterogeneity (e.g. Heckman 1981). This problem is addressed by formulating a separate equation for the mortality risk of the first-born child of every mother, which can be estimated jointly with equations (1) and (2):

$$(3) \quad M_{il}^* = g_l(\mathbf{x}_i, \mathbf{x}_{il}; \theta_{m,l}) + \lambda_m\alpha_{mi} + \lambda_b\alpha_{bi} + \lambda_f\alpha_{fi} + u_{mi1};$$

$$M_{il}=1 \text{ if } M_{il}^* > 0 \text{ and } M_{il}=0 \text{ if } M_{il}^* < 0$$

---

<sup>14</sup> Heckman, Hotz and Walker (1985) show, for a sample of Swedish mothers, that there is no state dependence in the birth spacing process once controls for unobserved heterogeneity are introduced.



In most existing applications of these sorts of models, described by Heckman (1981), Hyslop (1999) and Wooldridge (2000), the true process is ongoing and the first observation is generated in the same way as later observations, the only difference being that it is the first observation in the sampling window. Heckman, Hotz and Walker (1985) is an exception. They model birth spacing and observe the process from its natural start, the start of menarche. Here, similarly, we observe the birth and mortality processes from their beginning for each mother in the sample, and the first child is a genuine starting point of that process (as in Arulampalam and Bhalotra 2004a). This makes Heckman's approach quite natural compared to, for example, the alternative approach to addressing initial conditions recently proposed by Wooldridge (2000).

We will work with a linear specification of  $g_l$ , in line with the specification of (1). It seems likely that  $M_{il}$  will be correlated with  $\alpha_{mi}$  but since the equation for  $M_{il}$  contains no lagged dependent variable, the coefficient on  $\alpha_{mi}$  is allowed to be different from 1 (by  $\lambda_m$ ).  $M_{il}$  is also allowed to be correlated with  $\alpha_{bi}$  or  $\alpha_{fi}$ , the family-specific effects in the birth-spacing equation, (2), and the fertility equation, (4), introduced below. The error term  $u_{mil}$  is assumed to be standard normal and independent of the other error terms in the model, of the individual effects, and of the exogenous regressors  $x_{ij}$  and  $x_i$ .  $\theta_{m1}$ ,  $\lambda_m$ ,  $\lambda_b$  and  $\lambda_f$  are auxiliary parameters. Equation (3) is a flexible function of the exogenous variables. We do not impose restrictions on the relation of the parameters in (3) (risk for first born child) to those in (1) (risk for other children in the family).

### 3.4 Right Censoring

Inclusion of the birth spacing equation, (2), in the model demands a correction for right-censoring because some mothers will not have completed their fertility at the time of the survey.<sup>15</sup> The data contain information on whether a mother is sterilized at the time of the survey. For these mothers, who constitute 22.5% of the sample, it is safe to assume

---

<sup>15</sup> It may be useful to think in terms of the fertility equation being to the birth-spacing equation, what the participation equation is to the wage equation in the more familiar context of selection into wage work (e.g. Heckman 1974). The birth interval equation only applies if the woman has decided to have another child, i.e., if what we call the fertility equation has a certain (binary) outcome. The interval is infinity if the woman decides to have no more children. This is analogous to the wage being zero if the individual does not participate in market work.

that the complete birth process is observed. Of the remaining mothers, some will have another child after the survey date, and others will not. To account for this, we model the probability that mother  $i$  will have another child after the birth of child  $j$ , as follows:

$$(4) \quad F_{ij}^* = f(\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{ij}, M_{i1}, \dots, M_{ij-1}, B_{i2}, \dots, B_{ij}; \theta_j) + \alpha_{fi} + u_{fij};$$

$$F_{ij}=1 \text{ if } F_{ij}^* > 0 \text{ and } F_{ij}=0 \text{ if } F_{ij}^* < 0$$

We specify  $f$  as a linear combination of  $\mathbf{x}_i$ , the calendar year and age of the mother at the time of the birth of child  $j-1$  and their squares (functions of  $\mathbf{x}_{i1}$  and  $B_{i2}, \dots, B_{ij-1}$ ), dummies for the presence of boys and the presence of girls in the family (that did not suffer neonatal death), and the total numbers of boys and girls in the family who survived the neonatal period (functions of  $j$ ,  $M_{i1}, \dots, M_{ij}$ , and  $B_{i2}, \dots, B_{ij-1}$ ). The variables are gender specific to allow for son-preference, of which there is considerable evidence for the state of UP (e.g. Dreze and Gazdar 1997). Endogeneity of the gender-specific sibship variables is taken care of in the same way as in the other equations – these variables are a function of lagged dependent variables. Moreover, confounding unobserved factors are controlled for by allowing arbitrary correlations of  $\alpha_{fi}$  with  $\alpha_{mi}$  and  $\alpha_{bi}$ , assuming joint trivariate normality with arbitrary covariance matrix and independence of exogenous variables. We make similar assumptions on  $u_{fij}$  as on the other error terms: normality, independence of individual effects and error terms for other birth orders or other equations, and independence of exogenous variables.

Equation (4) is estimated jointly with equations (1)-(3). If mother  $i$  has more than  $j$  children, then we know she has given birth to another child after child  $j$ , and the likelihood will incorporate the probability that  $F_{ij}=1$ . If the mother reports that she has had exactly  $j$  children and was sterilized after the birth of the  $j$ -th child, then the likelihood will incorporate the probability that  $F_{ij}=0$ . If at the time of the survey, the mother had  $j$  children but was not (yet) sterilized, then it is unclear whether child  $j$  is the last child or not; it could be that the birth interval after the birth of child  $j$  extends beyond the time of the survey. The probability that this will happen, given that there will be another birth and given unobserved heterogeneity components, follows from (2) and is given by  $\Phi([T - \{h(\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{ij-1}, M_{i1}, \dots, M_{ij-1}, B_{i2}, \dots, B_{ij-1}; \theta_b) + \alpha_{bi}\}]/\sigma)$ , where  $T$  is the

length of the time interval elapsed between the birth of child  $j$  and the time of the survey, and  $\sigma$  is the standard deviation of the error term in (2). In this case, the likelihood (conditional on unobserved heterogeneity terms) will contain a factor that accounts for the fact that we do not observe whether or not there will be another birth after birth  $j$ :

$$\Phi(z_{ij}'\beta_c + \alpha_{fi}) \Phi([T - \{h(x_i, x_{i1}, x_{ij-1}, M_{i1}, \dots, M_{ij-1}, B_{i2}, \dots, B_{ij-1}; \theta_b) + \alpha_{bij}\}]/\sigma) + 1 - \Phi(z_{ij}'\beta_c + \alpha_{fi}).$$

The usual approach to right censoring is to assume that the same process continues but that we simply stop observing it at the time of the survey (e.g. Wooldridge, 2002, Chapter 20). This approach does not work well in the current application since the fertility process is necessarily finite (though at different points for different women) and ended well before the time of the survey for many women in the sample.<sup>16</sup> In the absence of information on sterilization, natural but less promising alternatives would be to assume that fertility stops at a given age (e.g. 40) for all mothers, or to estimate equation (4), but without the sterilization information. In this case, the fertility equation would only be indirectly identified in the sense that we would observe many women with very long birth intervals, and the model estimates would attribute this to cessation of fertility. These estimates are likely to be much less precise than we obtain with the sterilization information.

### 3.5 Geographical Cluster Effects

The data are collected in 333 geographical clusters (“communities”) with, on average, 21.3 mothers per cluster. This large number of clusters makes it infeasible to use cluster dummies. Instead, we incorporate random cluster effects in equations (1) and (2) and (4) in the same way as the mother specific effects, with similar assumptions (trivariate normal with arbitrary covariance structure to be estimated, independent of exogenous variables and error terms). A linear combination of the cluster effects in (1), (2) and (4) is added to equation (3), with three additional auxiliary parameters as coefficients. For identification, it is assumed that the geographical cluster effects are independent of mother specific effects. Thus common characteristics of all mothers in a

---

<sup>16</sup> Initial experimentation with our data showed that the usual procedure produces a poor fit, being unable to explain why so many women suddenly completely stop having children. This is because it merges the birth interval with the fertility decision, when in fact we need two separate equations for these two processes.

given community will be picked up by the cluster effects rather than by the mother specific effects.

### 3.6 Estimation

The complete model can be estimated by ML, including the nuisance parameters of the initial conditions equation, and the fertility equation determining whether or not there will be another birth. Conditional on the random (mother and cluster level) effects, the likelihood contribution of a given mother can be written as a product of univariate normal probabilities and densities over all births of a mother, and the likelihood for a given cluster can be written as the product over all mothers in that cluster. Since random effects are unobserved, the actual likelihood contribution is the expected value of the conditional likelihood contribution, with the expected value taken over all random effects (three in the model without cluster effects, six in the model with cluster effects). This is a three or six-dimensional integral, which could in principle be approximated numerically using, for example, the Gauss-Hermite-quadrature.

In this paper, we instead use (smooth) simulated ML, drawing multivariate errors from  $N(0, I_3)$ . These are then transformed into draws of the random effects using the parameters of the random effects distribution. The conditional likelihood contribution is then computed for each draw and the mean across  $R$  independent draws is taken. If  $R \rightarrow \infty$  with the number of observations (i.e., in this case, clusters, since mothers are no longer independent observations), this gives a consistent estimator; if draws are independent across households and  $R \rightarrow \infty$  faster than  $\sqrt{N}$ , then the estimator is asymptotically equivalent to exact ML (see, for example, Hajivassiliou and Ruud 1994). We use Halton draws, which have been shown to give more accurate results for smaller values of  $R$  than independent random draws (see Train 2003). The results we present are based on  $R=50$ . Using  $R=75$  gives very similar results (see section 4).

## 4 Results

This section first presents the results of the benchmark model (Tables 1-4) and then, in section 4.5, we discuss sensitivity of the results to some changes in specification

(Table 6). Table 5 presents the estimated covariance structure of the mother and community level random effects.

#### 4.1 Neonatal Mortality

Table 1 presents the parameter estimates of the equation for neonatal mortality. It also reports marginal effects for the second child, assuming that the first child survived the first month of life, and setting all family characteristics to their benchmark values when categorical (boy, Hindu, not of a backward caste, maternal and paternal education zero), and to their average values for second children when not (birth year 1985.7, age of the mother at birth 20.8 years, previous log birth interval 3.31).<sup>17</sup> The estimated probability of neonatal mortality for this benchmark child is 4.33%.

The preceding birth interval has the expected negative effect on the probability of neonatal death. A ten percent increase in the length of the birth interval reduces the probability of death by about 0.4 percentage-points in the benchmark case, and the marginal effect is similar for higher birth orders. To allow for non-linearity, we included the square of the log birth interval in the model, but this was insignificant. We also investigated a specification that is piecewise linear in the log birth interval (with four brackets given by the quartiles of the birth interval distribution) but, again, we were unable to reject the reported specification against this more general specification. This seems in line with the simple association shown in Figure 1. In view of the finding, in previous research, that the deleterious effects of short birth intervals are enhanced if the previous sibling has survived (e.g. Zenger 1993, Cleland and Sathar 1984), we also included an interaction of “lagged” neonatal mortality and the log of the preceding birth interval but this was insignificant. This interaction term is similarly insignificant in the analysis of data from Pakistan by Cleland and Sathar (1984), who interpret it as evidence that maternal depletion rather than sibling competition explains the mortality-increasing effects of short birth intervals

Neonatal mortality of the previous sibling makes neonatal death significantly more likely for the index child, even with the birth interval held constant. For the

---

<sup>17</sup> The marginal effects are birth-order-specific. A full set of marginal effects by birth order is available on request; not shown for parsimony.

benchmark second child, the estimated difference is 4.3 percentage-points. Similar effects are found for the third, fourth, and later children. This suggests that there are state dependence mechanisms other than those that operate via the birth interval (see section 1), and that these dominate any learning effects.

Conditional on the other covariates, neonatal mortality of boys and girls is not significantly different. Neonatal mortality is also not sensitive to birth order. Gender and birth order may influence later childhood mortality, with biological factors dominating neonatal mortality. For the benchmark child, there is a trend reduction of 0.15 percentage-points per year (1.9% of the benchmark probability) in the risk of death. Neonatal mortality risk is U-shaped in mother's age at birth of the index child, a pattern familiar from other studies using developing country data. The minimum occurs, in these data, at about 26 years of age. On average, mothers are much younger than this when giving birth to their second child (20.9 years old). This explains the significantly negative marginal effect obtained for the benchmark second child: if the mother's age increases by one year, the mortality probability falls by 0.17 percentage-points. At higher birth order, the average age of the mother increases and the U-shape implies that for birth orders above four, the marginal effect turns positive. For example, it is 0.16 percentage-points per year for the benchmark seventh child. Mortality risk tends to be decreasing in both maternal and paternal education, larger and more significant marginal effects being associated with maternal education. A striking result, that deserves further investigation, is that children of Muslim families are significantly less likely to die in the first month than Hindu children, with an estimated difference of about 1.6 percentage-points.<sup>18</sup> Although the scheduled castes and tribes face similar mortality risk to the benchmark case, other backward castes face risks of neonatal death that are higher by about 1.6 percentage points. These results are, again, of considerable interest, and merit further research.

---

<sup>18</sup> The raw data probability of neonatal and infant death is also lower amongst Muslims. This is puzzling given that Muslims, as compared with Hindus, exhibit shorter birth intervals, higher fertility and a greater proportion of mothers and fathers with no education. Note that the state of UP (for which data are analysed in this paper) has, at 17%, an above average representation of Muslims in the population.

Estimates of the “reduced form” probit equation for mortality of *first-born* children (equation 3) are in Table 2. The female dummy is now negative and significant at the two-sided 10% level, consistent with previous research that shows that discrimination against girls is increasing in birth order (e.g. DasGupta 1990). Other effects are broadly similar.

## 4.2 Birth Spacing

Estimates of the birth spacing equation are in Table 3. Since the dependent variable is in logs, the interpretation of the parameters is in terms of percentage changes of the expected length of the birth interval. Note that all covariates in this model refer to the preceding child (i.e. the child born at the start of the birth interval).

There is a strong negative effect of neonatal death of the previous child on the subsequent birth interval, reducing its expected length by about 20.5%. This is consistent with replacement behaviour (e.g. Ben-Porath 1976).<sup>19</sup>

The gender of the last-born child is significant, consistent with son-preference. If this was a girl, the expected birth interval is about 3% shorter than if it was a boy. The quadratic trend is hump shaped, with a maximum at about 1978. Thus birth intervals have tended to get shorter in recent decades (1978-1998). This may be explained by rising living standards. In particular, since better-nourished mothers will tend to suffer less deleterious effects from a short birth interval, they can “afford” shorter birth intervals. There is some indication that spatial (inter-state) patterns in India resemble the inter-temporal pattern detected here, with the wealthier states (like Punjab) having a greater proportion of births with short intervals while, at the same time, having lower neonatal mortality (see Arulampalam and Bhalotra 2004b). Birth spacing is hump-shaped in the age of the mother at birth, with a maximum at about 28 years of age. This means that, for the average mother, birth intervals increase until the sixth child is born. Parental education has no significant effect on birth spacing. Birth intervals are shorter amongst Muslim families by 8%, compared with Hindu families. There are no significant

---

<sup>19</sup>Hoarding in view of expected mortality will, in general, result in a positive correlation of the unobserved heterogeneity terms in the mortality and fertility equations. This is not so relevant in the current context since mortality is defined as neonatal. In this case, parents have (neonatal) mortality information on all previous children before they decide to have the next child.

differences in birth spacing by ethnic group. Birth-order exhibits a non-monotonic pattern, with the shortest birth intervals preceding the birth of the third and fourth child.

### 4.3 Fertility Equation

Table 4 presents estimates of the probability of having another child after each birth, as a function of current family composition, other family characteristics and calendar time. Of particular interest are the family composition variables. The results indicate son-preference, of which there is considerable evidence from Northern India and, especially, the state of UP (e.g. Dreze and Sen 1997). The probability of continued fertility is decreasing in the number of surviving children, but more rapidly in the number of surviving boys. Also, if the family has no surviving boys, the probability of having another child is much larger than if there are no surviving girls. Similar results have been reported for other countries in Asia and North Africa (e.g. Choe *et al* 1998, Nyarko *et al* 2003)<sup>20</sup>

The quadratic in the child's year of birth is hump-shaped, with a maximum at about 1979. So, for the latter two decades of the data, fertility has been declining. The quadratic in mother's age is U-shaped, with a minimum at about 31 years. In the sample, 89% of births were to mothers younger than this so, for most cases, (conditional) fertility is falling in maternal age. Continued fertility is seen to fall with the level of education of both mother and father, larger effects of a given level of education being associated with mothers. Muslims and families of other non-Hindu religions show higher tendencies to continue fertility, as do all of the backward castes.

### 4.4 Unobserved Heterogeneity

Table 5 presents the estimated covariance structure of the mother and community level random effects. From now onwards the sum of these effects is referred to as *total* unobserved heterogeneity. The underlying auxiliary parameters are presented in the bottom panel of the table. There is evidence of mother and community specific effects in all equations. Compared to the idiosyncratic noise term (with variance 1), the two



heterogeneity terms in the mortality equation make a modest contribution, capturing about one seventh of the total unsystematic variation in  $M_{ij}^*$  ( $0.1675/(1+0.1675)$ ). Most of this is heterogeneity across communities, only about 20% of it is across mothers within communities. Previous research in rich and poor countries has found evidence of mother-level frailty, with varying estimates of its contribution to the overall variation in mortality risk (e.g. Rosenzweig and Schultz 1983a,b, Curtis *et al* 1993, Guo 1993, Zenger 1993) but these studies typically do not allow for clustering at the community level and so they will tend to over-estimate the mother effects (see Sastry 1997b, Bolstad and Manda 2001, Nyarko *et al* 2003).

In the equation for the log birth interval, the idiosyncratic noise term has estimated variance 0.204 ( $0.452^2$ ), and the heterogeneity terms together pick up only about 10% of the total unsystematic variation. We find no evidence of correlation between either the mother specific or the community specific heterogeneity terms in the birth spacing and neonatal mortality equations: correlations between community and mother specific effects in the two equations are insignificant (the parameters  $\pi_{bm}$  and  $\tau_{bm}$  in the bottom panel of the table). Moreover, the estimated covariances are of opposite sign and almost cancel out against each other, giving a correlation coefficient of  $-0.004$  for the total unobserved heterogeneity terms.

The heterogeneity terms in the fertility equation explain about 16% of the unsystematic variation in  $F_{ij}^*$ . Correlation with the heterogeneity terms in the mortality equation is insignificant, but we find a large negative and significant correlation between mother specific effects in the fertility and birth interval equations of  $-0.92$ , inducing a negative correlation between the total unobserved heterogeneity terms of  $-0.44$ . These estimates indicate that mothers with a desire to have many children also tend to have shorter birth intervals, keeping observed explanatory variables constant. This is consistent with, for example, the target fertility model (see Heckman, Hotz and Walker 1985, Wolpin 1997).

Overall, the heterogeneity terms are statistically significant but relatively small compared to the idiosyncratic errors. This raises the question of whether neglecting to

---

<sup>20</sup> Angrist and Evans (1998) find no such asymmetry for the US; they do find that the probability of a third child is larger if the first two children are of the same sex than if they are of different

allow for unobserved heterogeneity would lead to biased estimates of the parameters of interest. This question is explored in the following section.

Table 5 also shows how the unobserved heterogeneity terms enter the equation for neonatal mortality of the first child. As expected, mothers with a relatively large probability of neonatal mortality of higher birth order children are also more likely to experience higher mortality risk for the first child. Somewhat surprisingly, we do not find the same for the community effects - these are insignificant in the mortality equation for the first child. We also find that mothers with a tendency towards higher fertility face a larger probability of neonatal mortality for the first child (the significantly positive value of  $\pi_{0f}$ ).

#### 4.5 Sensitivity Analysis

Table 6 presents estimates of the coefficients on the (lagged) endogenous variables for alternative specifications. The effects of the other variables are not shown as they do not change much compared to the benchmark model (Model 1), estimates of which are in Tables 1-5. Consider the consequences of omitting the birth interval from the mortality equation (Model 2). This increases the estimated effect of lagged mortality in the mortality equation, consistent with the mechanisms described in section 1, whereby previous mortality causes a shortening of the birth interval and this, in turn, leads to an increase in mortality risk for the subsequent child. Omission of the birth interval from equation (1) also biases the effect of lagged mortality on the birth interval in equation (2). The reason is that according to the Model 2 estimates, there is a significant and substantial negative correlation (of  $-0.49$ ) between the (total) unobserved heterogeneity terms in equations (1) and (2), and this creates an upward simultaneity adjustment on the coefficient of lagged mortality. This differs from Model 1 with its small and insignificant correlation (of  $-0.004$ ) between total unobserved heterogeneity terms in equations (1) and (2) (see section 4.4).

Model 3 excludes the community effects. The estimated covariance matrix of the mother specific effects in this model is similar to the covariance matrix of the sum of mother and community specific effects in the complete model, with, for example, a very

small correlation between the terms in the mortality and birth interval equations (0.014). This explains why the point estimates are very similar to those in the benchmark model.<sup>21</sup> The main difference is that this model underestimates the standard errors on account of its ignoring correlations across observations. These results are in line with those of Sastry (1997b).

Model 4 does not allow unobserved heterogeneity at the community or mother level. In spite of the modest role of the mother specific heterogeneity terms that we saw in the benchmark model (Table 5), failure to control for heterogeneity creates some significant changes. The most salient one is the effect of lagged mortality on current mortality, which is 77% larger in Model 4 than in the benchmark model (and 66% larger than in the model that allows mother-specific but not community-specific unobserved heterogeneity: model 3). This is in line with the traditional argument that ignoring heterogeneity leads to overestimation of state dependence effects (Heckman 1981). There is little change in the effect of mortality on the next birth interval, probably because the correlation between the total unobserved heterogeneity terms is very close to zero in the benchmark model.

Model 5 combines the restrictions imposed in arriving at Models 2 and 4. The two positive biases on the effect of lagged mortality on mortality together lead to an estimate that is almost 95% larger than in the benchmark model. There is no bias on the coefficient of mortality in the birth interval equation, for the same reason as in Model 4.

A challenging finding is that the effect of lagged mortality on mortality remains so strong even when the length of the preceding birth interval is controlled for. An example of a mechanism that may result in state dependence effects without involving changes in birth interval length is maternal depression, but there could well be other mechanisms at play. Here, we consider the possibility that the significance of lagged mortality in equation (1) reflects a specification error. For example, the family may have suffered a temporary shock (a poor harvest, maternal illness) that spans two or more

---

<sup>21</sup> There are some changes in significance of the other covariates. For example, the paternal education terms become significant in the mortality equation, whereas, in the benchmark case, which adjusts for geographic clustering, only the maternal education terms are significant. Similarly, Model 3 shows significant effects of maternal education on birth spacing which, in the benchmark case, are insignificant.

births, resulting in greater vulnerability of two successive children. This was investigated by including the second lag of the neonatal mortality dummy in equation (1) (Model 6, Table 6). The coefficient on the second lag is positive and statistically significant (0.246 with standard error 0.062). Instead of reducing the effect of the first lag (as would be expected if  $M_{ij-2}$  were in fact an omitted variable),<sup>22</sup> it limits the role of unobserved heterogeneity: the standard deviation of the unobserved heterogeneity term in the mortality equation falls from 0.38 to 0.24 (standard error 0.062). Thus it seems that the results suggested by our benchmark model cannot be put down to misspecification of the lag structure.<sup>23,24</sup>

We now discuss some additional sensitivity checks that were conducted, results for which are not presented in Table 6 since they were virtually identical to those of the benchmark model. The results presented so far are based on 50 random draws for each observation ( $R=50$ ). Extending this to 75 draws hardly changes the results.<sup>25</sup> We found higher neonatal mortality amongst children with a shorter preceding birth interval. In order to ensure that this is not simply the result of a selective over-representation of premature births (as noted by Eastman 1944, cited in Cleland and Sathar 1984, p406), we re-estimated the model after removing from the sample all mothers with at least one birth interval under 9 months. This resulted in a loss of 40 mothers (0.6% of all mothers). The estimates of the main parameters are virtually the same as when the short birth intervals are included.

---

<sup>22</sup> It may be better to compare the state dependence estimate of 0.482 to the estimate of the same coefficient in a model without second lag but with a separate equation for mortality of the second child. Such a model gives a coefficient of 0.419 (with standard error 0.069).

<sup>23</sup> An alternative would be to allow for autocorrelation between the error terms in the mortality equation. We experimented with this in a single equation framework (using the GHK algorithm to obtain the simulated likelihood) but found an insignificant (negative) autocorrelation coefficient rather than the positive coefficient that would be expected under the hypothesis that the significance of  $M_{ij-1}$  reflects a temporary shock. Also, there is, again, an increase in the coefficient on the lagged dependent variable.

<sup>24</sup> For computational convenience and given the similarity of the results for models 1 and 3, we did not incorporate community clusters in this variant of the model. To do this would require specification of a separate equation not only for the first but also for the second child (for whom the second lag cannot be included). See Heckman (1981); details available upon request.

<sup>25</sup> The effect of lagged mortality on mortality is 0.337 (s.e. 0.066), the effect of the log birth interval on mortality is -0.486 (s.e. 0.048), and the effect of mortality on the log birth interval is -0.230 (s.e. 0.018).

As already mentioned above, adding an interaction term of the log birth interval and lagged mortality in the mortality equation does not lead to a significant improvement. The interaction term has coefficient 0.058 with t-value 0.47. At the median log birth interval value (3.258), this gives a coefficient 0.360 on the birth interval, similar to the benchmark model value. The coefficient on the log birth interval is -0.491 if the previous child did not die, similar to the benchmark value of -0.481. The estimates of the birth interval equation are virtually identical to those in the benchmark model. Similar minimal deviations compared to the benchmark model are found when the square of the log birth interval is added to the mortality equation. The point estimate on this is 0.013 with t-value 0.27.

## 5 Conclusions

Using retrospective fertility histories from a large sample of Indian mothers, a dynamic panel data model is estimated that describes the complete process of child survival and birth spacing (and thus also fertility), allowing for endowment heterogeneity, input endogeneity, right-censoring and the initial conditions problem. It offers the first rigorous estimates of the causal effect of mortality on subsequent birth spacing, and of the extent to which death clustering amongst siblings can be explained by endogenously determined short birth intervals.

We find evidence that childhood mortality risk is influenced by the pattern of childbearing, that is, by the timing and spacing of births, and that birth-spacing and continued fertility are, in turn, a function of realized mortality. The latter is consistent with replacement behaviour and, in our data, these estimates indicate son-preference. Together, these recursive causal effects suggest multiplier effects of policies that reduce mortality or lengthen birth intervals.

Our results show that unobserved heterogeneity in the form of mother or community specific effects explains part of the correlation between neonatal mortality of successive children observed in the data. Another part is explained through birth spacing: neonatal mortality shortens the birth interval to the next child, and this increases the risk of another neonatal death. The largest part of the correlation, however, is explained by neither the birth interval mechanism nor unobserved heterogeneity and could, for

example, be due to a mental health shock induced by the death of a child, leading to maternal behaviour that increases the chances of subsequent mortality. This is a striking result, especially as previous demographic research has restricted attention to the birth spacing mechanism.

We find clear evidence of mother specific heterogeneity (“frailty” in the mortality equation and “fecundity” in the birth spacing equation) and of community specific effects. This underlines the importance of the joint panel data framework that is used in this paper. Using data on sterilization to estimate an equation for the decision to have another child at each birth, we find that women who want to have many children also tend to choose shorter birth intervals, a result that has some intuitive appeal.

Future work could extend the framework to analyze infant or child (under-5) mortality. This creates the additional complication that mortality events and births can take place in overlapping time periods, requiring a different modeling approach. Other extensions could make explicit use of data on breastfeeding, although this would mean restricting the analysis to recent births as these data are not available in most DHS surveys for children born more than five years before the survey. Finally, these results are for one Indian state, albeit a state with a population estimated at more than 166 million in 2001 (compared with a population of about 59 million in the UK!). Extension of the analysis to consider other Indian states or other developing countries will lend important insight into the extent to which the key relationships analysed here are altered by socio-economic development.

## References

- Akerlof, G. (1998), Men without children, *The Economic Journal*, Vol. 108, No. 447, 287-309.
- Alessie, R., S. Hochguertel and A. van Soest (2004), Ownership of stocks and mutual funds: a panel data analysis, *Review of Economics and Statistics*, Vol. 86, No. 3, 783-796.
- Angrist, J.D. and Evans, B (1998). “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size,” *American Economic Review*, Vol. 88, No. 3, 450-477.
- Arulampalam, W. and Bhalotra, S. (2004a), Sibling death clustering in India: genuine scarring vs unobserved heterogeneity. Working Paper 03/552, Department of Economics, University of Bristol. Revised February 2004.

Arulampalan, W. and S. Bhalotra (2004b), Infant survival in India: frailty and state-dependence, Working Paper 04/558, Department of Economics, University of Bristol, May.

Ben-Porath, Y. (1976), Fertility response to child mortality: Micro Data from Israel, *Journal of Political Economy*, 84(4): S163-S178.

Bhargava, Alok (2003), Family planning, gender differences and infant mortality: evidence from Uttar Pradesh, India, *Journal of Econometrics*, 112, 225-240.

Bolstad, W.M. and Manda, Sam O. (2001), Investigating child mortality in Malawi using family and community random effects: A Bayesian analysis, *Journal of the American Statistical Association*, Vol. 96, 12-19.

Bongaarts, J. and Potter, R.G. (1983), *Fertility, Biology and Behaviour: An Analysis of the Proximate Determinants*. New York: Academic Press.

Chen, L., S. Ahmed, M. Gesche, and W. Mosley (1974), A prospective study of birth interval dynamics in Rural Bangladesh, *Population Studies*, Vol. 28, 277-297.

Cherlin, A. (1990), Recent changes in american fertility, marriage, and divorce, *Annals of the American Academy of Political and Social Science*, Vol. 510, July, 145-154.

Choe, M.K., I. Diamond, F. Steele and S.K. Kim (1998), Son preference, family building process and child mortality, in United Nations, *Too Young to Die: Genes or Gender*, UN Secretariat, Population Division, New York, 270-294.

Cigno, A (1998), "Fertility decisions when infant survival is endogenous", *Journal of Population Economics*, Vol. 11, No. 1, 21-28

Cleland, J. and Z. A. Sathar (1984), The effect of birth spacing on childhood mortality in Pakistan, *Population Studies*, Vol. 38, 401-418.

Curtis, S. L., Diamond and I., and McDonald J. W. (1993), Birth interval and family effects on postneonatal mortality in Brazil , *Demography*, Vol. 33, No. 1, 33-43.

DasGupta (1990), "Death Clustering, mothers education and the determinants of child-mortality in rural Punjab, India", *Population Studies*, Vol. 44, No. 3, 489-505.

DaVanzo, Julie and Anne R. Pebody (1993), Maternal depletion and child survival in Guatemala and Malaysia, Labor and Population Program Working Paper 93-18, RAND.

Drèze, J., and Gazdar, H. (1997) "Uttar Pradesh: The Burden of Inertia", in Drèze, J., and Sen, A. (eds.) *Indian Development: Selected Regional Perspectives*, Oxford University Press, 33-128.

Dreze, Jean and Amartya Sen (1997), *Indian Development: Selected Regional Perspectives*, Oxford: Clarendon Press.

Eastman, NJ (1944) "The Effect of the Interval Between Births on A Maternal and Fetal Outlook," *American Journal of Obstetrics and Gynecology*, Vol. 47, 445.

Frankenberg, E. (1998), The relationship between infant and child mortality and subsequent fertility in Indonesia, 1971-1991, in M. Montgomery and B. Cohen (eds.), *From Death to Birth: Mortality Decline and Reproductive Change*, National Research Council, National Academy Press, Washington DC: 254-315.

Galor, O., Weil, D. (2000), Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and beyond, *The American Economic Review*, Vol. 90, No. 4, Sept., 806-828.

Gribble, J.N. (1993), Birth intervals, gestational age and low birth weight: are the relationships confounded?, *Population Studies*, Vol. 47, 133-146.

Griliches, Z. (1979), Sibling Models and Data in Economics: Beginnings of a Survey," *Journal of Political Economy*, Vol. 87, No. 5, Part 2, S37-S64.

Guo, Guang (1993), Use of sibling data to estimate family mortality effects in Guatemala, *Demography*, Vol. 30, No. 1, 15-32.

Hajivassiliou, V. and P. Ruud (1994), "Classical estimation methods for LDV models using simulation," in: R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, North-Holland, New York, 2384-2443.

Heckman, James J (1974) Shadow prices, market wages, and labor supply, *Econometrica*, vol. 42(4), pages 679-94, July.

Heckman, J.J. (1981), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, in C.F. Manski and D.L. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, London: MIT Press, 179-195.

Heckman, J., V.J. Hotz and J.Walker (1985), New evidence on the timing and spacing of births, *American Economic Review*, 75: 179-84.

Hyslop, D.R. (1999), State dependence, serial correlation and heterogeneity in inter-temporal labor force participation of married women, *Econometrica*, 67, 1255-1294.

IIPS and ORC Macro (2000), *National Family Health Survey (NFHS-2) 1998-9: India*. Mumbai: International Institute for Population Sciences (IIPS).

Madise, N.J. and I. Diamond (1995), Determinants of infant mortality in Malawi: an analysis to control for death clustering within families. *Journal of Biosocial Science*, Vol. 27, No. 1, 95-106.

Manski, C (1993) Identification of Endogenous Social Effects: The Reflection Problem, *Review of Economic Studies*, Vol. 60, 531-542.

Mattheissen, P. and McCann, J. 1978. "The Role of Mortality in the European Fertility Transition: Aggregate-level Relations", in: S.H. Preston, ed., *The Effects of Infant and Child Mortality on Fertility*. Academic Press, New York, USA.



Moffitt, R. (2003), Causal analysis in population research: an economist's perspective, Mimeograph, Johns Hopkins University, July.

Nyarko, P., Madise, N.J. and Diamond, I (2003) "Child loss and fertility behaviour in Ghana". Social Statistics Research Centre Working Paper, A03/08, University of Southampton.

Olsen, R. (1980), Estimating the effect of child mortality on the number of births, *Demography*, Vol. 17(4), November, 429-443.

Olsen, R. and K. Wolpin (1983), The impact of exogenous child mortality on fertility: a waiting time regression with exogenous regressors, *Econometrica*, Vol. 51, 731-749.

Preston, S. H. (1985), Mortality in childhood: lessons from WFS. In: Cleland, J.G. and Hobcraft, J., eds. Reproductive Change in Developing Countries. Oxford: Oxford University Press, 46-59.

Rosenzweig, M. and T.P. Schultz (1983a), Consumer demand and household production: the relationship between fertility and child mortality, *American Economic Review*, Vol. 73, 38-42.

Rosenzweig, M. and T.P. Schultz (1983b), Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight, *The Journal of Political Economy*, Vol. 91, No. 5, 723-746.

Rosenzweig, M. and T.P. Schultz (1989), Schooling, information and nonmarket productivity: Contraceptive use and its effectiveness, *International Economic Review*, Vol. 30, No. 2, 457-477.

Rosenzweig, M. and K. Wolpin (1988), Heterogeneity, intrafamily distribution and child health, *The Journal of Human Resources*, Vol. 23, No. 4, 437-461.

Rosenzweig, M. and K. Wolpin (1995), Sisters, siblings and mothers: the effect of teenage childbearing on birth outcomes in a dynamic family context, *Econometrica*, Vol. 63, No. 2, 303-326.

Sastry, Narayan 1997a, Family-level clustering of childhood mortality risk in Northeast Brazil, *Population Studies*, Vol. 51, No. 3, 245-261.

Sastry, Narayan 1997b, A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil, *Journal of the American Statistical Association*, Vol. 92, No. 438, June, 426-435.

Scrimshaw, S. (1996), Nutrition and health from womb to tomb, *Nutrition Today*, 31, 55-67.

Solon, G., Corcoran, M. Gordon, R. and Laren, D. (1991), A longitudinal analysis of sibling correlations in economic-status, *Journal of Human resources*, Vol. 26, No. 3, 509-534.

Steer, R. A., Scholl, T.O., Hediger, M. L. and Fischer, R. L. (1992), Self-reported depression and negative pregnancy outcomes, *Journal of Clinical Epidemiology*, Vol. 45, No. 10, 1093-1099.

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

UNDP (2003), Human Development Report: *Millenium Development Goals: A Compact Among Nations To End Human Poverty*.

Wolpin, Kenneth (1997), Determinants and consequences of the mortality and health of infants and children, Chapter 10 in Mark Rosenzweig and Oded Stark (eds.), *Handbook of Population and Family Economics*, Volume 1A, Amsterdam: North Holland.

Wooldridge, J. (2000), A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables, *Economics Letters*, Vol.6, No. 3, 245-250.

Wooldridge, J. (2002), *Econometric Analysis of Cross-section and Panel Data*, MIT Press, Cambridge (Mass.).

World Bank (2004), Attaining the Millennium Development Goals in India: How likely and what will it take to reduce infant mortality, child malnutrition, gender disparities and hunger-poverty and to increase school enrolment and completion?, Human Development, South Asia, April.

Zenger, E. (1993), Siblings' neonatal mortality risks and birth spacing in Bangladesh. *Demography*, Vol. 30, No. 3, 477-488.

**Table 1: Neonatal Mortality (equation 1)**

	Parameter	Std error	ME	Std.err(ME)
lagged mortality	0.346*	0.067	4.321	1.173
log birth interval	-0.481*	0.050	-4.404	0.576
<b>Religion:</b>				
Muslim	-0.201*	0.068	-1.553	0.482
Other	-0.083	0.363	0.431	3.941
<b>Ethnicity:</b>				
scheduled caste	0.057	0.056	0.560	0.528
scheduled tribe	0.166	0.129	1.964	1.595
other backward caste	0.149*	0.045	1.589	0.494
<b>Maternal education:</b>				
incomplete primary	-0.092	0.100	-0.690	0.768
complete primary	-0.184	0.095	-1.383	0.684
incomplete secondary	-0.132	0.100	-1.033	0.761
secondary & higher	-0.313*	0.144	-2.063	0.840
<b>Paternal education:</b>				
incomplete primary	-0.002	0.091	0.050	0.848
complete primary	-0.097	0.072	-0.810	0.600
incomplete secondary	-0.114	0.060	-0.981	0.493
complete secondary	-0.134	0.074	-1.112	0.579
higher than secondary	-0.016	0.066	-0.141	0.612
<b>Gender:</b>				
Female	-0.039	0.038	-0.340	0.320
<b>Trend effects:</b>				
year of birth of child/10	0.232	0.675	2.805	6.503
(year/10) squared	-0.023	0.040	-0.252	0.383
<b>Maternal age:</b>				
maternal age at birth/10	-0.849*	0.350	-8.067	3.233
(age/10) squared	0.161*	0.064	1.522	0.583
			-0.172	0.100
<b>Child birth order</b>				
birth order	0.042	0.055	0.383	0.498
square of birthorder	-0.002	0.005	-0.016	0.046
Constant	0.574	2.870		

Notes: \*: parameter (and marginal effect) significant at the two-sided 5% level. ME denotes marginal effects. These are computed for a benchmark child, defined in section 4.1. The (omitted) reference cases for the categorical variables (religion, ethnicity, maternal and paternal education, gender) are defined in Appendix Table 1.

**Table 2. Neonatal Mortality of the first-born child (equation 3)**

	Parameter	Std error	t-value
<b>Religion:</b>			
Muslim	-0.272	0.217	-1.25
Other	-0.039	0.843	-0.05
<b>Ethnicity:</b>			
scheduled caste	0.070	0.180	0.39
scheduled tribe	0.457	0.437	1.05
other backward caste	0.331	0.180	1.84
<b>Maternal education:</b>			
incomplete primary	0.197	0.283	0.70
complete primary	-0.186	0.279	-0.67
incomplete secondary	0.039	0.264	0.15
secondary & higher	-0.489	0.319	-1.53
<b>Paternal education:</b>			
incomplete primary	-0.429	0.316	-1.36
complete primary	0.043	0.220	0.19
incomplete secondary	-0.162	0.190	-0.85
complete secondary	-0.162	0.219	-0.74
higher than secondary	-0.288	0.234	-1.23
<b>Gender:</b>			
female	-0.262	0.149	-1.76
<b>Trend effects:</b>			
year of birth of child/10	-1.137	1.627	-0.70
(year/10) squared	0.055	0.098	0.56
<b>Maternal age:</b>			
maternal age at birth/10	-2.977	1.986	-1.50
(age/10) squared	0.588	0.484	1.21
constant	6.178	7.024	0.88

See Notes to Table 1.

**Table 3. Log Birth Interval (equation 2)**

	Parameter	Std error	t-value
<b><i>lagged mortality</i></b>	-0.230*	0.017	-13.40
<b>Religion:</b>			
Muslim	-0.080*	0.014	-5.81
Other	-0.109	0.065	-1.68
<b>Ethnicity:</b>			
scheduled caste	0.003	0.013	0.25
scheduled tribe	0.039	0.033	1.17
other backward caste	0.006	0.010	0.61
<b>Maternal education:</b>			
Incomplete primary	0.008	0.024	0.31
complete primary	0.035	0.021	1.69
Incomplete secondary	0.013	0.024	0.53
secondary & higher	0.033	0.023	1.48
<b>Paternal education:</b>			
Incomplete primary	0.023	0.021	1.10
complete primary	-0.001	0.016	-0.06
Incomplete secondary	0.002	0.015	0.16
complete secondary	-0.003	0.017	-0.19
higher than secondary	0.017	0.018	0.94
<b>Gender:</b>			
Female	-0.029*	0.009	-3.38
<b>Trend effects:</b>			
year of birth of child/10	1.082*	0.138	7.82
(year/10) squared	-0.069*	0.008	-8.35
<b>Maternal age:</b>			
maternal age at birth/10	0.265*	0.070	3.76
(age/10) squared	-0.047*	0.015	-3.15
<b>Child birth order</b>			
birth order	-0.024*	0.008	-2.93
square of birthorder	0.003*	0.001	2.72
Constant	-1.130*	0.570	-1.98
<i>sigma error</i>	0.452*	0.003	164.85

Notes: See Notes to Table 1.

**Table 4. Fertility: Probability of another birth (equation 4)**

	Parameter	Std error	ME	Std.err(ME)
<b>Religion:</b>				
Muslim	1.223*	0.111	22.201	1.711
Other	-0.628	0.334	-23.419	12.371
<b>Ethnicity:</b>				
scheduled caste	0.209*	0.060	6.203	1.779
scheduled tribe	0.329*	0.132	9.191	3.405
other backward caste	0.299*	0.049	8.548	1.416
<b>Maternal education:</b>				
incomplete primary	-0.213	0.110	-7.309	3.852
complete primary	-0.164	0.087	-5.696	3.025
incomplete secondary	-0.298*	0.093	-10.824	3.544
secondary & higher	-0.520*	0.105	-19.111	3.990
<b>Paternal education:</b>				
incomplete primary	0.232*	0.097	6.768	2.547
complete primary	-0.133	0.082	-4.496	2.779
incomplete secondary	-0.192*	0.060	-6.458	2.063
complete secondary	-0.448*	0.079	-15.883	2.979
higher than secondary	-0.362*	0.085	-12.597	3.126
<b>Gender:</b>				
female	5.520*	1.200	177.148	40.961
<b>Trend effects:</b>				
year of birth of child/10	-0.351*	0.070	-11.260	2.391
(year/10) squared			-2.439	0.261
<b>Maternal age:</b>				
maternal age at birth/10	-3.164*	0.412	-100.263	14.044
(age/10) squared	0.515*	0.073	16.287	2.463
			0.012	0.277
<b>Surviving children</b>				
1 if no boys	1.073*	0.173	34.303	5.847
1 if no girls	0.220*	0.075	7.017	2.358
number of boys	-0.329*	0.029	-10.497	1.080
number of girls	-0.069*	0.024	-2.234	0.781
constant	-14.584*	4.996		

Notes: See Notes to Table 1. The marginal effects are given for the benchmark case after the birth of the seventh child, with probability almost 75% of having another child. The reason for taking the 7<sup>th</sup> child is that probabilities of having another child after an earlier birth are larger, giving rather small marginal effects.

**Table 5. Unobserved Heterogeneity**

<b>Mother-specific effects:</b>			
Covariance matrix			
	Mortality	Birth interval	Fertility
Mortality	0.032		
Birth interval	0.003	0.019	
Fertility	0.016	-0.039	0.096
Correlation matrix			
	Mortality	Birth interval	Fertility
Mortality	1.000		
Birth interval	0.111	1.000	
Fertility	0.286	-0.921	1.000
<b>Community-specific effects:</b>			
Covariance matrix			
	Mortality	Birth interval	Fertility
Mortality	0.136		
Birth interval	-0.003	0.003	
Fertility	-0.031	0.015	0.089
Correlation matrix			
	Mortality	Birth interval	Fertility
Mortality	1.000		
Birth interval	-0.144	1.000	
Fertility	-0.285	0.689	1.000
<b>Mother plus community effects:</b>			
Covariance matrix			
	Mortality	Birth interval	Fertility
Mortality	0.167		
Birth interval	0.000	0.022	
Fertility	-0.016	-0.028	0.185
Correlation matrix			
	Mortality	Birth interval	Fertility
Mortality	1.000		
Birth interval	-0.004	1.000	
Fertility	-0.089	-0.437	1.000

<b>Table 5 continued</b>			
<b>Underlying parameter estimates:</b>			
	Parameter	Std error	t-value
$\pi_{mm}$	0.178*	0.054	3.32
$\pi_{bm}$	0.015	0.017	0.91
$\pi_{bb}$	0.137*	0.012	11.93
$\pi_{fm}$	0.089	0.090	0.99
$\pi_{fb}$	-0.297*	0.074	-4.02
$\pi_{ff}$	-0.005	0.140	-0.04
$\pi_{0m}$	1.755*	0.624	2.81
$\pi_{0b}$	-0.205	0.196	-1.04
$\pi_{0f}$	0.995*	0.360	2.76
$\tau_{mm}$	0.369*	0.046	8.06
$\tau_{bm}$	-0.008	0.014	-0.56
$\tau_{bb}$	0.055*	0.024	2.32
$\tau_{fm}$	-0.085	0.082	-1.04
$\tau_{fb}$	0.195	0.150	1.30
$\tau_{ff}$	0.209	0.108	1.94
$\tau_{0m}$	0.077	0.188	0.41
$\tau_{0b}$	-0.007	0.252	-0.03
$\tau_{0f}$	-0.223	0.189	-1.18

Notes: Refer section 4.4 of the text.

\*: parameter (and marginal effect) significant at the two-sided 5% level

Mother specific effects are parameterized as follows:

Mortality:  $\alpha_{mi} = \pi_{mm}u_{mi}$ ;

Birth interval:  $\alpha_{bi} = \pi_{bm}u_{mi} + \pi_{bb}u_{bi}$ ;

Fertility:  $\alpha_{fi} = \pi_{fm}u_{mi} + \pi_{fb}u_{bi} + \pi_{ff}u_{fi}$ ;

$u_{mi}$ ,  $u_{bi}$ ,  $u_{fi}$  independent standard normal, independent of exogenous variables and error terms. The parameters  $\pi_{0m}$ ,  $\pi_{0b}$ , and  $\pi_{0f}$  are the coefficients of  $u_{mi}$ ,  $u_{bi}$ , and  $u_{fi}$  in the equation for neonatal mortality of the first child.

Community specific effects are parameterized as follows:

Mortality:  $(_{mi} = \tau_{mm}v_{mi}$ ;

Birth interval:  $(_{bi} = \tau_{bm}v_{mi} + \tau_{bb}v_{bi}$ ;

Fertility:  $(_{fi} = \tau_{fm}v_{mi} + \tau_{fb}v_{bi} + \tau_{ff}v_{fi}$ ;

$v_{mi}$ ,  $v_{bi}$ ,  $v_{fi}$  independent standard normal, independent of  $u_{mi}$ ,  $u_{bi}$ ,  $u_{fi}$ , exogenous variables, and error terms. The parameters  $\tau_{0m}$ ,  $\tau_{0b}$ , and  $\tau_{0f}$  are the coefficients of  $v_{mi}$ ,  $v_{bi}$ , and  $v_{fi}$  in the equation for neonatal mortality of the first child.



**Table 6. Sensitivity Analysis:  
Endogenous coefficients in alternative specifications**

	<b>Mortality equation</b>		<b>Birth interval eq.</b>
	<b><math>M_{i,j-1}</math></b>	<b><math>\ln B_{i,j}</math></b>	<b><math>M_{i,j-1}</math></b>
<b>Model 1</b>			
Benchmark model	0.346 (0.067)	-0.481 (0.050)	-0.230 (0.017)
<b>Model 2</b>			
No lagged birth interval	0.444 (0.067)		-0.198 (0.017)
<b>Model 3</b>			
No cluster effects	0.369 (0.055)	-0.480 (0.039)	-0.234 (0.014)
<b>Model 4</b>			
No unobserved heterogeneity	0.614 (0.036)	-0.444 (0.034)	-0.236 (0.012)
<b>Model 5</b>			
No lagged birth interval & no unobserved heterogeneity	0.673 (0.036)		-0.236 (0.012)
<b>Model 6</b>			
2 <sup>nd</sup> lag in mortality equation & no cluster effects	0.482 (0.068)	-0.505 (0.044)	-0.234 (0.014)

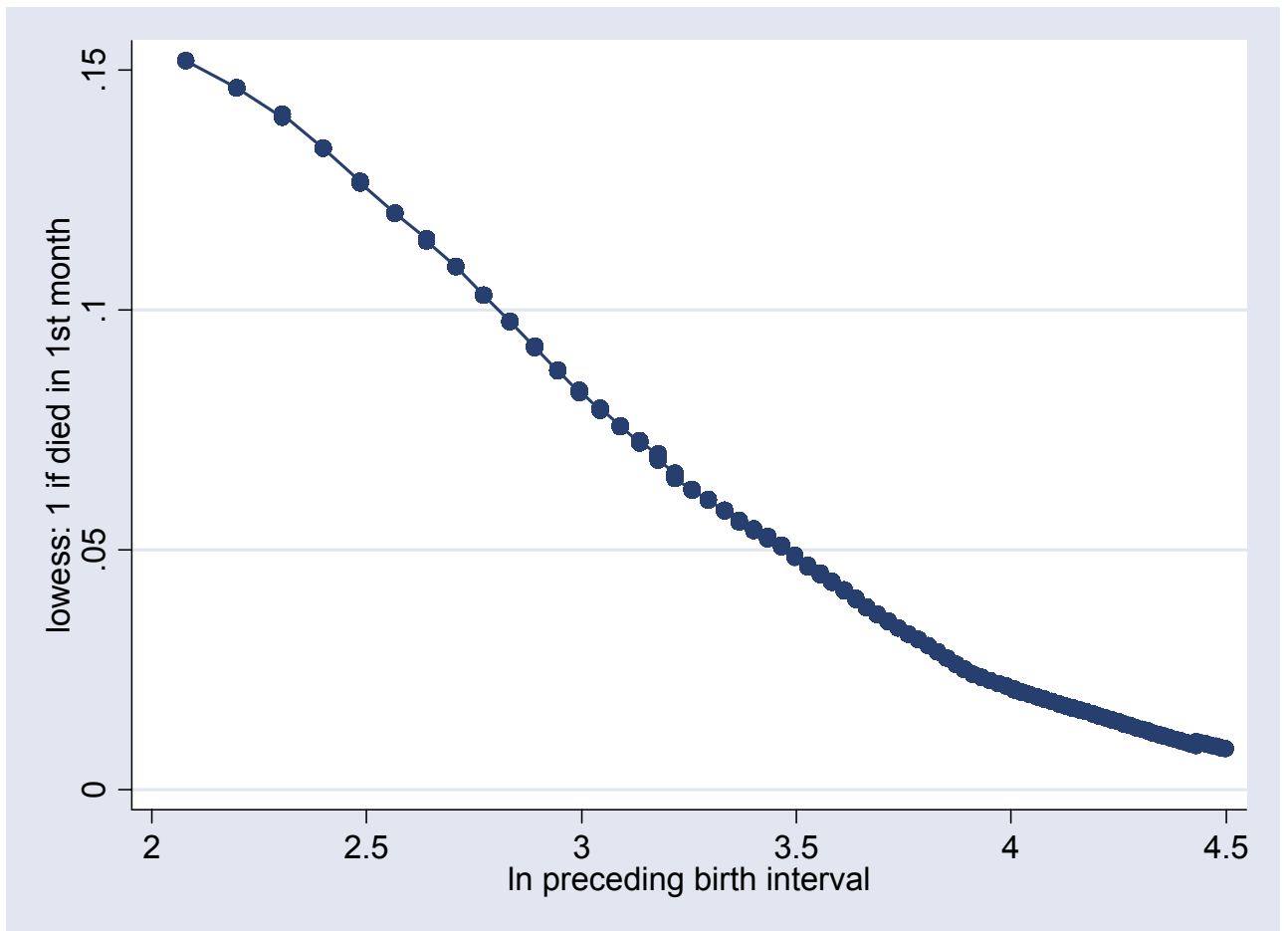
Notes: See section 4.5 of the text for explanation. Figures in parentheses are standard errors.

**Appendix Table 1:**  
**Variable Definitions and Summary Statistics**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
neonatal mortality	0.070		0.0	1.0
lagged neonatal mortality	0.060		0.0	1.0
log birth interval*	3.293	0.484	2.1	5.7
<i>Hindu</i>	0.825		0.0	1.0
Muslim	0.167		0.0	1.0
other religions	0.007		0.0	1.0
<i>not backward caste</i>	0.454		0.0	1.0
scheduled caste	0.196		0.0	1.0
scheduled tribe	0.021		0.0	1.0
other backward caste	0.276		0.0	1.0
<i>mother has no education</i>	0.755		0.0	1.0
ma has incomplete primary	0.045		0.0	1.0
ma has completed primary	0.076		0.0	1.0
ma has incomplete secondary	0.060		0.0	1.0
ma has secondary or higher	0.064		0.0	1.0
<i>father has no education</i>	0.335		0.0	1.0
pa has incomplete primary	0.068		0.0	1.0
pa has completed primary	0.111		0.0	1.0
pa has incomplete secondary	0.193		0.0	1.0
pa has completed secondary	0.125		0.0	1.0
pa has higher than secondary	0.164		0.0	1.0
female	0.475		0.0	1.0
year of birth of child*	86.510	7.173	63.0	98.0
maternal age at birth*	23.140	5.507	12.0	47.0
birth order*	3.160	2.040	1.0	14.0
no surviving boys	0.122		0.0	1.0
no surviving girls	0.188		0.0	1.0
number of surviving boys*	1.962	1.386	0.0	8.0
number of surviving girls*	1.782	1.461	0.0	10.0

Notes: All variables other than those with a \* are dummies. Lagged mortality refers to the mortality status of the preceding sibling. Italics indicate reference category omitted in the regressions. The number of children is 28668 except for log birth interval, for which it is 21567. The number of mothers is 6716.

**Figure 1**  
**Nonparametric (lowess) relation of (predicted) neonatal mortality and**  
**the preceding birth interval**



Notes: The top 1% of observations were deleted.

**Figure 2**

**Density of log birth interval by survival status of preceding sibling**

